

A Retrospective Analysis Comparing the New Standardized Letter of Recommendation in Dermatology with the Classic Narrative Letter of Recommendation

^aJESSICA A. KAFFENBERGER, MD; ^bJOY MOSSER, MD; ^aGRACE LEE, MD; ^aLLANA POOTRAKUL, MD, PhD; ^bKATYA HARFMANN, MD; ^aSTEPHANIE FABBRO, MD; ^bESTEBAN FERNANDEZ FAITH, MD; ^aDAVID CARR, MD; ^aALISHA PLOTNER, MD; ^aMATTHEW ZIRWAS, MD; ^aBENJAMIN H. KAFFENBERGER, MD

^aDepartment of Internal Medicine, Division of Dermatology, The Ohio State University Wexner Medical Center, Columbus, Ohio;

^bDepartment of Pediatrics, Division of Pediatric Dermatology, The Ohio State University and Nationwide Children's Hospital, Columbus, Ohio

ABSTRACT

Background: In an effort to avoid numerous problems associated with narrative letters of recommendation, a dermatology standardized letter of recommendation was utilized in the 2014–2015 resident application cycle. **Objective:** A comparison of the standardized letter of recommendation and narrative letters of recommendation from a single institution and application cycle to determine if the standardized letter of recommendation met its original goals of efficiency, applicant stratification, and validity. **Methods:** Eight dermatologists assessed all standardized letters of recommendation/narrative letters of recommendation pairs received during the 2014–2015 application cycle. Five readers repeated the analysis two months later. Each letter of recommendation was evaluated based on a seven question survey. Letter analysis and survey completion for each letter was timed. **Results:** Compared to the narrative letters of recommendation, the standardized letter of recommendation is easier to interpret ($p < 0.0001$), has less exaggeration of applicants' positive traits ($p < 0.001$), and has higher inter-rater and intrarater reliability for determining applicant traits including personality, reliability, work-ethic, and global score. Standardized letters of recommendation are also faster to interpret ($p < 0.0001$) and provide more information about the writer's background or writer-applicant relationship than narrative letters of recommendation ($p < 0.001$). **Limitations:** This study was completed at a single institution. **Conclusions:** The standardized letter of recommendation appears to be meeting its initial goals of 1) efficiency, 2) applicant stratification, and 3) validity. (*J Clin Aesthet Dermatol.* 2016;9(9):36–42.)

Letters of recommendation (LOR) are an important component of the dermatology residency application and can provide valuable information regarding personality, reliability, and work ethic, which are not explicitly addressed elsewhere in the application. Unfortunately, despite their importance in resident selection, LOR may be difficult to interpret^{1,2} and several studies question their utility.^{1,3–7} Narrative letters of recommendation (NLOR) are often excessively flattering^{3–5} with a complicated hierarchy of laudatory phrases⁴ or “code words”/expressions.^{8,9} Additionally, they can be redundant, simply reciting other parts of the application.² Finally, NLOR have been found to lack clarity^{1,5} and have demonstrated low

reliability between interpreting faculty members.⁶

In an effort to improve the LOR process, the authors created a standardized letter of recommendation (SLOR), which was accepted by the Association of Professors of Dermatology (APD) for use along with the traditional narrative letter of recommendation in the 2014–2015 application cycle. The dermatology SLOR is based on currently used models of SLOR from emergency medicine, otolaryngology, plastic surgery, and orthopedic surgery. The dermatology SLOR also incorporates qualities that are deemed important by members of the APD,² such as the letter writer's background and specific applicant traits (e.g., personality, work ethic and reliability). The SLOR strives to

DISCLOSURE: This article has no funding source. The authors report no relevant conflicts of interest. IRB from The Ohio State University Wexner Medical Center was approved for this study.

ADDRESS CORRESPONDENCE TO: Jessica Kaffenberger, MD, Division of Dermatology, Department of Internal Medicine Ohio State University, 540 Officenter Place, Suite 240, Gahanna, OH 43230; E-mail: jessica.kaffenberger@osumc.edu

improve the NLOR in the following three primary areas: 1) efficiency—decreased time and effort needed for interpretation, 2) stratification—improved ability to discern the quality of an applicant, and 3) validity—high consistency in inter-rater interpretation and intrarater reliability. Since the SLOR is a new addition to the dermatology application process, it is essential to evaluate if it has met its initial goals. The primary aim of this study is to compare the SLOR and NLOR to determine if the SLOR meets these three criteria. A secondary goal is to assess the amount of information provided by each type of LOR.

METHODS

Study design. This was a retrospective examination of letters of recommendation from the 2014–2015 cycle. All letters of recommendation received by the Ohio State Division of Dermatology for the 2014–2015 cycle were included provided there was at least one narrative and one standardized letter for the same applicant. All letters were de-identified by a physician not involved in the interpretation and the pair of SLOR and NLOR was assigned a number. There were 46 pairs of SLORs/NLORs, however two pairs written by the same author were excluded due to excessively generic narrative letters (serving only as a supplement). Six dermatology faculty members and two dermatology chief residents completed the LOR interpretations. Each interpreter was assigned a number to ensure anonymity. The interpreter evaluated consecutively all letters of one type before evaluating the corresponding letters in the other format. The order of evaluation was randomly assigned.

Each LOR was evaluated based on a seven question survey, which was developed based on the goals of the SLOR and previous literature^{10,11} (see Supplements 1 and 2). Each reader evaluated the LOR by determining how strongly he/she would rank the applicant based on the LOR, the difficulty or ease in determining the strength of the LOR, the amount of time and effort required to evaluate the letter, and the perceived accuracy versus exaggeration of the letter. All survey questions were based on a seven rung Likert scale as utilized in previous literature.^{10,11} Letter analysis and survey completion was timed for each letter.

To determine intrarater reliability, five readers repeated the complete analysis of all LORs two months after completing the original evaluation.

Statistical analysis. *Time and ease of interpretation.*

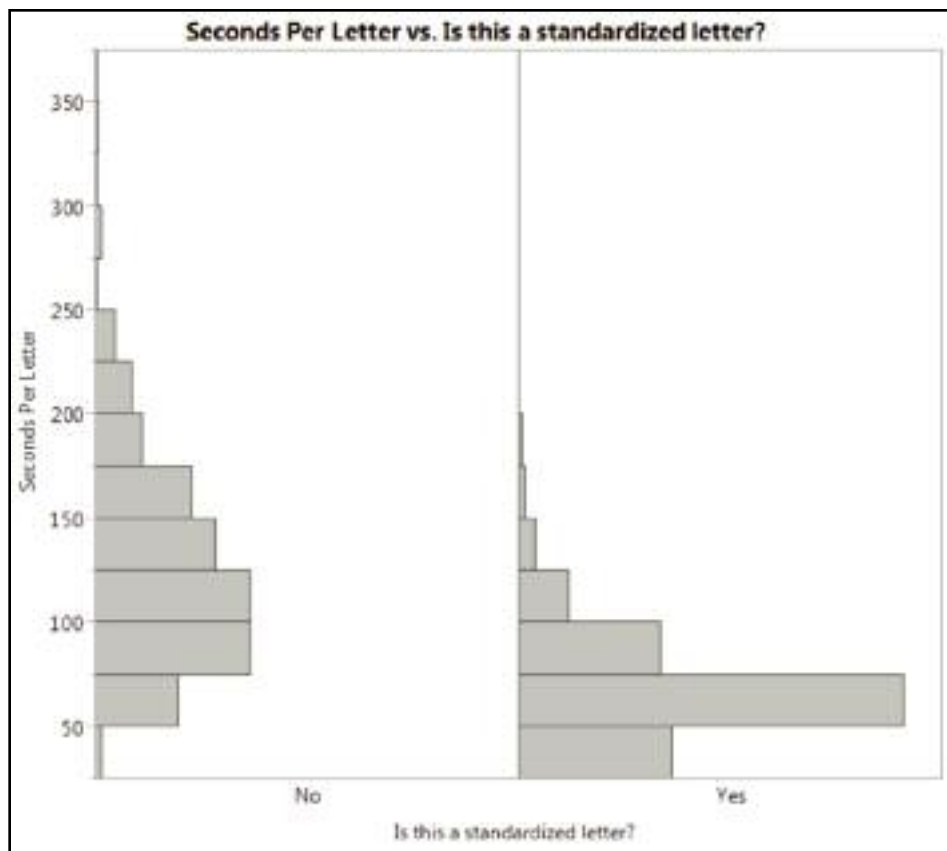


Figure 1. Seconds per letter between standardized and narrative letters. A mean 128 seconds was spent evaluating narrative letters compared to 68 seconds to evaluate the standardized letter ($p < 0.0001$).

This was performed by consecutively subtracting timestamps at the completion of each letter. The interpreters were given strict instructions to complete letters consecutively and if a stoppage was needed, to make a false entry to restart the timestamps. The initial letter was thus not evaluable. Distributions were then analyzed for each letter/interpreter combination. Outliers based on time of completion were removed from the data to prevent unavoidable distractions from affecting the overall results. The times to interpret and answer letters were then analyzed based on using a paired two-tailed t -test. A Spearman ρ was performed to evaluate the subjective effort to interpret letter with the actual time involved. A χ^2 analysis was performed to evaluate differences in ordinal values in determining strength of recommendation, overall ease of letter interpretation, and perception of letter inflation.

Reliability analysis. Inter-rater reliability was analyzed using the Kendall coefficient of concordance calculated with ties (W). These were calculated separately for each question. Intrarater reliability was analyzed using the Spearman rank correlation coefficient (ρ) between interpretations and the mean of the 5 was reported.

RESULTS

Time and ease of interpretation. There was a statistically reliable difference in time to interpret the

TABLE 1. Reliability analysis

	NARRATIVE LETTER	STANDARDIZED LETTER
Personality		
Inter-rater - W	0.437	0.654
Intrarater - Averaged ρ	0.368	0.754
Work Ethic		
Inter-rater - W	0.481	0.720
Intrarater - Averaged ρ	0.542	0.906
Reliability		
Inter-rater - W	0.456	0.723
Intrarater - Averaged ρ	0.511	0.882
Perception of Letter Inflation		
Inter-rater - W	0.282	0.364
Intrarater - Averaged ρ	0.426	0.507
Global Score		
Inter-rater - W	0.578	0.704
Intrarater - Averaged ρ	0.643	0.764
W=Kendall Coefficient of Concordance ρ =Spearman Rank Correlation Coefficient (averaged among 5 repeat interpreters)		

corresponding letters (Figure 1). On average, NLOR required more than two minutes to interpret compared to one minute for the corresponding SLOR ($p < 0.0001$). Subjective effort to complete the interpretations had a linear correlation with the actual time spent in evaluation and was significant both by analysis of variance (ANOVA) and a Spearman $\rho = 0.688$, $p < 0.0001$. A χ^2 analysis indicated the SLOR was more easily evaluated in comparison to the NLOR ($p < 0.0001$). Similarly, the SLOR was more easily interpreted from a global standpoint ($p < 0.0001$), and the applicants' positive traits were felt to be less exaggerated ($p < 0.0001$).

Reliability analysis. The inter-rater reliability testing was generally higher within standardized letters of recommendation compared to the narrative letters by the Kendall W and similar findings were present when analyzing the intrarater reliability utilizing the averaged Spearman ρ (Table 1).

Information provided. There was a statistically reliable difference in amount of information regarding the writer's background or the relationship of the writer to the applicant. On average, SLOR had 4.4 pieces of information while NLOR

had 2.3 pieces of information regarding the writer's background or writer-applicant relationship ($p < 0.0001$).

DISCUSSION

Letters of recommendation are deemed an important part of the application process^{1,12-14}; however, research suggests numerous problems are associated with the NLORs. In light of these NLOR weaknesses, nearly 80 percent of a group of academic dermatologists were in support or possibly in support of developing a SLOR.²

The authors' study indicates that the new dermatology SLOR is significantly more efficient than the NLOR. Each fall, dermatology programs are inundated with hundreds of applications. Saving over a minute of time per each letter has the potential to save evaluators hours of time. This mirrors the efficacy demonstrated in evaluating the emergency medicine (EM) SLOR¹⁰ and otolaryngology SLOR.¹¹ Additionally, although not addressed in the authors' study, research has shown that SLOR also have a shorter composition time compared to NLOR,¹¹ which would further improve the efficiency of the application process.

The authors' study suggests that reviewers deem applicants positive traits to be less exaggerated with the NLOR than the SLOR. Academic dermatologists are often unwilling to address an applicants' negative qualities,² and previous research has found that NLOR contain only positive feedback about an applicant.⁸ Therefore, the authors postulate that NLOR may seem more laudatory because the writer dwells only on the applicants' positive traits, while weaknesses are ignored. In contrast, the SLOR forces the writer to evaluate the entire applicant. However, the authors understand the potential to exaggerate an applicant's positive traits on a SLOR still exists and research on the EM SLOR demonstrates that grade inflation is a problem.^{15,16} Interestingly, the EM task force suggests that grade inflation could be minimized if the word limit on the small narrative section of the EM SLOR was followed.¹⁵ The goal of the written section on the EM SLOR is to concisely address any superlative or negative ratings in the application. Unfortunately, with nearly half of EM SLOR having written comments over the 200-word-limit, the current status of this section falls prey to the same problems of NLOR.¹⁵ Given this research, the dermatology SLOR was structured to include only a small written section; however, the impact of this section on letter interpretation is yet to be determined.

The authors also found that SLOR have higher inter-rater and intrarater reliability than NLOR, which is consistent with previous research.^{10,11} NLOR have worse reliability likely because of varying interpretations of select phrases and "code words" (e.g., "excellent" versus "outstanding" versus "superb").^{8,9} Additionally, research has demonstrated significant differences between female and male letter writers. Female letter writers are more likely to focus on the applicant's compassion and ability to work as part of a team.⁸ The authors also postulate that readers get fatigued after reading large numbers of lengthy NLOR and this could affect letter interpretation.

Lastly, previous research indicates that the background

SUPPLEMENT 1

REVIEWER NUMBER _____

LETTER NUMBER _____

IS THIS A STANDARDIZED LETTER? Y OR N

PLEASE CIRCLE THE NUMBER THAT BEST CORRESPONDS TO YOUR ANSWER FOR THE FOLLOWING QUESTIONS:

1) How many pieces of information regarding the writer's background and/or relationship with the applicant is provided?

1	1 or less pieces of information
2	2 pieces of information
3	3 pieces of information
4	4 pieces of information
5	5 or more pieces of information

2) Based on this letter of recommendation please rank the applicant on his/her:**A) PERSONALITY**

0	No information on personality provided
1	Undesirable personality
2	
3	Personality below average
4	
5	Personality above average
6	
7	Best personality

B) WORK ETHIC

0	No information on work ethic provided
1	Poor work ethic
2	
3	Work ethic below average
4	
5	Work ethic above average
6	
7	Best work ethic

C) RELIABILITY

0	No information on reliability provided
1	Most un reliable
2	
3	More un reliable than average
4	
5	More reliable than average
6	
7	Most reliable

Continued on next page

SUPPLEMENT 1. Cont.**3) How easy is it to determine how strongly the letter writer is recommending this applicant?**

1	Very difficult to determine how strongly the applicant was being recommended
2	
3	Somewhat difficult to determine how strongly the applicant was being recommended
4	
5	Somewhat easy to determine how strongly the applicant was being recommended
6	
7	Very easy to determine how strongly the applicant was being recommended

4) How much time and effort did it take to evaluate this letter?

1	Minimal amount of time needed to interpret the letter
2	
3	Less than average amount of time needed to interpret the letter
4	
5	More than average amount of time needed to interpret the letter
6	
7	Extensive amount of time needed to interpret the letter

5) How much exaggeration of the applicant's positive traits is present in the letter?

1	No exaggeration of applicant's traits/ No excessive flattery
2	
3	Less than average exaggeration of applicant's traits
4	
5	More than average exaggeration of applicant's traits
6	
7	Extensive exaggeration of applicant's traits/excessive flattery

6) Based on this letter of recommendation, please globally rank this applicant.

SCORE	RANKING
1	Very poor, would not rank
2	
3	Good, would rank mid-lower
4	
5	Excellent, would rank mid-upper
6	
7	Superb, would rank guaranteed to match

of the letter writer as well as the length and type of relationship between the writer and applicant is extremely important in evaluating LOR.² Surprisingly, NLOR contain only an average of 2.3 pieces of information regarding the background of the letter writer or about the writer-applicant relationship. It is plausible that writers who do not know the applicant well forgo including information about the writer-

applicant connection so attention is not drawn to the weak relationship. In contrast, the SLOR's structured questions provide the reader more information about the writer-applicant relationship and writer's background, which provides a better framework to understand the recommendation.

Limitations. The authors' study has some limitations.

SUPPLEMENT 2. Protocol

A COMPARISON OF A NEW STANDARDIZED LETTER OF RECOMMENDATION IN DERMATOLOGY WITH THE CLASSIC NARRATIVE LETTER OF RECOMMENDATION

JESSICA A. KAFFENBERGER; MATTHEW ZIRWAS, MD; BENJAMIN H. KAFFENBERGER, MD

I. OBJECTIVES

To compare a newly developed dermatology standard letter of recommendation with the traditional narrative letter of recommendation with regard to its:

- A) time of interpretation and thus ease of use
- B) ability to stratify applicants
- C) validity as indicated by consistency in rater interpretation and inter-rater reliability

II. BACKGROUND AND RATIONALE

Letters of recommendation (LOR) are an important component of the dermatology residency application and can provide valuable information regarding personality, reliability, and work ethic, which are not explicitly addressed elsewhere in the application. Unfortunately, despite their importance in resident selection, LOR may be difficult to interpret¹ and several studies question their utility.¹⁻⁶ Narrative letters of recommendation (NLOR) are often excessively flattering²⁻⁴ with a complicated hierarchy of laudatory phrases³ or code words. Thus, it is not surprising that NLOR often lack clarity^{1,4} and have demonstrated low reliability between interpreting faculty members.^{5,7}

In the fall of 2013, we developed a standardized letter of dermatology (SLOR) that was accepted by the Association of Professors of Dermatology (APD) for use along with the traditional narrative letter of recommendation in the 2014–2015 application cycle. It is essential to analyze this new standardized letter of recommendation to ensure that it is meeting its goals of 1) decreased time of interpretation, 2) ability to stratify applicants, and 3) validity as indicated by consistency in rater interpretation and inter-rater reliability.

III. PROCEDURES

A. RESEARCH DESIGN

This is a retrospective examination of letters of recommendation from the 2014–2015 cycle. We will compile all letters of recommendation received by the dermatology department for the 2014–2015 cycle. All letters will be deidentified by a physician not involved in the analysis and the pair of SLOR and NLOR will be assigned a number. Each physician who evaluates the letter will also be assigned a number to ensure anonymity. All letters of one type will be evaluated before evaluating letters in the other format. The order of evaluation will be randomly assigned. Each pair will be evaluated based on a seven question survey. All survey questions will be based on a seven rung Likert scale. Letter analysis and survey completion was timed for each letter. A group of reviewers will repeat the complete analysis of the LOR two months following the evaluation to determine intrarater reliability.

B. SAMPLE

We will evaluate all SLOR and NLOR received in the department of dermatology for the 2014–2015 application cycle. We estimate there will be about 40–80 pairs of SLORs and NLORs.

C. MEASUREMENT / INSTRUMENTATION

A Likert scale from 1–7 will be utilized to evaluate each SLOR and NLOR. A rating of 1 would correspond to an extremely poor candidate while a 7 would correspond to an exceptional candidate. Dermatology faculty members will also indicate the ease of determining how strongly a candidate was being recommended (difficult to determine, moderately difficult to determine, moderately easy to determine, easy to determine) and will indicate the time it took them to evaluate each letter.

D. DETAILED STUDY PROCEDURES

All data will be viewed under a locked door at the OSU office facility and saved within the secured access dermatology research drive. Inter-rater reliability will be analyzed using the Kendall coefficient of concordance calculated with ties (W). Intrarater reliability will be analyzed using the Spearman rank correlation coefficient (ρ). Time and ease of interpretation will be analyzed based on using a paired two-tailed t -test.

E. INTERNAL VALIDITY

Utilizing a Likert scale from 1–9, we will determine inter-rater reliability and consistency of rater interpretation of SLOR and NLOR. External threats include that this is a single center academic facility and the results may not be generalizable to all dermatology academic centers.

Continued on next page

F. DATA ANALYSIS

Pearson's correlation coefficient and Cohen's Kappa coefficient will be used to determine inter-rater reliability and consistency in rater interruption.

IV. BIBLIOGRAPHY

1. DeZee KJ, Thomas MR, Mintz M, Durning SJ. Letters of recommendation: rating, writing, and reading by clerkship directors of internal medicine. *Teach Learn Med.* 2009;21:153–158. doi:10.1080/10401330902791347.
2. Fortune JB. The content and value of letters of recommendation in the resident candidate evaluative process. *Curr Surg.* 2002;59:79–83.
3. Morgenstern BZ, Zalneraitis E, Slavin S. Improving the letter of recommendation for pediatric residency applicants: an idea whose time has come? *J Pediatr.* 2003;143:143–144.
4. Keim S, Rein J, Chisholm C, et al. A standardized letter of recommendation for residency application. *Acad Emerg Med.* 1999;6:1141–1146.
5. Dirschl DR, Adams GL. Reliability in evaluating letters of recommendation. *Acad Med.* 2000;75:1029.
6. Shultz K, Mahabir RC, Song J, Verheyden CN. Evaluation of the current perspectives on letters of recommendation for residency applicants among plastic surgery program directors. *Plast Surg Int.* 2012;2012:728981. doi:10.1155/2012/728981.
7. Greenburg AG, Doyle J, McClure D. Letters of recommendation for surgical residencies: what they say and what they mean. *J Surg Res.* 1994;56:192–198.
8. Girzadas D V, Harwood RC, Dearie J, Garrett S. A comparison of standardized and narrative letters of recommendation. *Acad Emerg Med.* 1998;5:1101–1104.

First, this study was conducted at a single institution with a small number of faculty and therefore may not represent all dermatology programs. Secondly, this study incorporates data from highly experienced letter readers (more than 10 years of experience) to less experienced readers (first year of experience). Although this diversity of experience may mimic the reality in many programs, it may affect how much time was spent on the letters and how they were analyzed. Finally, a limitation of the SLOR itself is that many programs may be reluctant to implement this format. Because the SLOR requires evaluation of numerous qualities, it may hinder programs' weaker applicants.

CONCLUSION

Based on this study, the SLOR appears to be meeting its initial goals of 1) efficiency, 2) ability to stratify applicants, and 3) validity. However, given that this is a small study, larger studies will be needed to further analyze the SLOR as well as the impact it has on the dermatology application process.

REFERENCES

1. DeZee KJ, Thomas MR, Mintz M, Durning SJ. Letters of recommendation: rating, writing, and reading by clerkship directors of internal medicine. *Teach Learn Med.* 2009;21:153–158.
2. Kaffenberger BH, Kaffenberger JA, Zirwas MJ. Academic dermatologists' views on the value of residency letters of recommendation. *J Am Acad Dermatol.* 2014;71:395–396.
3. Fortune JB. The content and value of letters of recommendation in the resident candidate evaluative process. *Curr Surg.* 2002;59:79–83.
4. Morgenstern BZ, Zalneraitis E, Slavin S. Improving the letter of recommendation for pediatric residency applicants: an idea whose time has come? *J Pediatr.* 2003;143:143–144.
5. Keim S, Rein J, Chisholm C, et al. A standardized letter of recommendation for residency application. *Acad Emerg Med.* 1999;6:1141–1146.
6. Dirschl DR, Adams GL. Reliability in evaluating letters of recommendation. *Acad Med.* 2000;75:1029.
7. Shultz K, Mahabir RC, Song J, Verheyden CN. Evaluation of the current perspectives on letters of recommendation for residency applicants among plastic surgery program directors. *Plast Surg Int.* 2012;2012:728981.
8. Messner AH, Shimahara E. Letters of recommendation to an otolaryngology/head and neck surgery residency program: their function and the role of gender. *Laryngoscope.* 2008;118:1335–1344.
9. Greenburg AG, Doyle J, McClure D. Letters of recommendation for surgical residencies: what they say and what they mean. *J Surg Res.* 1994;56:192–198.
10. Girzadas D, Harwood R, Dearie J, Garrett S. A comparison of standardized and narrative letters of recommendation. *Acad Emerg Med.* 1998;5:1101–1104.
11. Perkins JN, Liang C, McFann K, et al. Standardized letter of recommendation for otolaryngology residency selection. *Laryngoscope.* 2013;123:123–133.
12. Bernstein AD, Jazrawi LM, Elbeshbeshy B, et al. An analysis of orthopaedic residency selection criteria. *Bull Hosp Jt Dis.* 2003;61:49–57.
13. Egol KA, Collins J, Zuckerman JD. Success in orthopaedic training: resident selection and predictors of quality performance. *J Am Acad Orthop Surg.* 2011;19:72–80.
14. Nguyen AT, Janis JE. Resident selection protocols in plastic surgery: a national survey of plastic surgery independent program directors. *Plast Reconstr Surg.* 2012;130:459–469.
15. Love JN, Deiorio NM, Ronan-Bentle S, et al. Characterization of the council of emergency medicine residency directors' standardized letter of recommendation in 2011–2012. *Acad Emerg Med.* 2013;20:926–932.
16. Harwood R, Girzadas D, Carlson A, et al. Characteristics of the Emergency Medicine Standardized Letter of Recommendation. *Acad Emerg Med.* 2000;7:409–410. ●